Indigenous health program evaluation design and methods in Australia: a systematic review of the evidence

Kamalini Lokuge,[1] Katherine Thurber,[1] Bianca Calabria,[1] Meg Davis,[2] Kathryn McMahon,[2] Lauren Sartor,[2] Raymond Lovett,[1] Jill Guthrie,[1] Emily Banks[1,3]

Abstract

Objective: Indigenous Australians experience a disproportionately higher burden of disease compared to non-Indigenous Australians. High-quality evaluation of Indigenous health programs is required to inform health and health services improvement. We aimed to quantify methodological and other characteristics of Australian Indigenous health program evaluations published in the peer-reviewed literature.

Methods: Systematic review of peer-reviewed literature (November 2009-2014) on Indigenous health program evaluation.

Results: We identified 118 papers describing evaluations of 109 interventions; 72.0% were university/research institution-led. 82.2% of evaluations included a quantitative component; 49.2% utilised quantitative data only and 33.1% used both quantitative and qualitative data. The most common design was a before/after comparison (30.5%, n=36/118). 7.6% of studies (n=9/118) used an experimental design: six individual-level and three cluster-randomised controlled trials. 56.8% (67/118) reported on service delivery/process outcomes (versus health or health risk factor outcomes) only.

Conclusions: Given the number of Indigenous health programs that are implemented, few evaluations overall are published in the peer-reviewed literature and, of these, few use optimal methodologies such as mixed methods and experimental design.

Implications for public health: Multiple strategies are required to increase high-quality, accessible evaluation in Indigenous health, including supporting stronger research-policy-practice partnerships and capacity building for evaluation by health services and government.

Key words: Aboriginal and Torres Strait Islander health, Indigenous health, program evaluation, evaluation methods

It is established that Aboriginal and Torres Strait Islander (hereafter referred to as Indigenous) people in Australia and Indigenous people internationally experience a disproportionately higher burden of disease than non-Indigenous people. In recognition of this, improving Indigenous health outcomes is a priority area for Australian policy and programs. High quality and appropriate evaluation must be an integral component of the program cycle if Indigenous health policy and practice is to evolve and improve based on lessons from past experiences. Further, these evaluations must be accessible to practitioners and policymakers through publication.

To our knowledge, there has been no systematic assessment of the quality, methods, outputs and utility of evaluations published across the Indigenous health sector.[1] This analysis therefore aimed to describe study designs utilised in current evaluations of Indigenous health interventions published in the peer-reviewed literature, and examine variation in design by type of evaluator.

Methods

Study inclusion criteria

Peer-reviewed publications were eligible for inclusion in this review if their primary aim was to describe the implementation and/or evaluation of services, programs or policies (hereafter referred to as 'interventions') aimed at improving the health of Indigenous Australians, either directly or indirectly. We limited our review to the five year period preceding the point of data collection as our aim was to describe current evaluation practice. In this study, we considered interventions to include any 'systematic actions and approaches taken to address an identified Indigenous health need'.[2] Given that Indigenous perspectives of health are holistic, we included studies related to physical, mental, social and emotional wellbeing. Study protocols of proposed evaluations were excluded, given that these interventions had not yet been evaluated.

1. National Centre for Epidemiology and Population Health, Research School of Population Health, The Australian National University, Australian Capital Territory
2. Medical School, The Australian National University, Australian Capital Territory
3. Sax Institute, New South Wales
Correspondence to: Dr Kamalini Lokuge, National Centre for Epidemiology and Population Health, Research School of Population Health, The Australian National University, 62 Mills Road, Acton, ACT 2601; e-mail: kamalini.lokuge@anu.edu
Submitted: December 2016; Revision requested: March 2017; Accepted: May 2017
The authors have stated they have no conflict of interest.

Aust NZ J Public Health. 2017; 41:480-2; doi: 10.1111/1753-6405.12704

Studies had to meet all three of the following eligibility criteria for inclusion.

1. The primary target population was Indigenous Australians.
2. The intervention's primary aim was to improve health and/or health services.
3. The study used primary information to describe the implementation and/or evaluation of an intervention (documents relying on secondary sources of information and/or data, including reviews, were excluded).

Search strategy

We identified peer-reviewed papers using LItSearch, a tool designed to improve the effectiveness of searches for Indigenous health literature from the PubMed database.[3,4] We used the additional filter: 'impact OR eval* OR implement*'.

Applying exclusion criteria

One assessor (author KT) appraised each identified document for inclusion against the above criteria on the basis of the information available in the document's title and abstract. If eligibility could not be determined based on these fields, the full text was reviewed. If a document failed to meet any of the three criteria, it was excluded from the study. If the search identified multiple papers published on the same intervention, we included those papers that referred to a discrete aspect of the evaluation (i.e. measuring different health outcomes or the same health outcome in a different target population, or evaluating impact versus process).

Data extraction

A group of four assessors (authors MD, KM, LS, KT) extracted data on the included studies (reviewing between 28 and 33 studies each). The full text of each included manuscript was reviewed to extract data on characteristics of the intervention, the process of implementing the intervention, and expected or measured outcomes of the intervention.[5-7] One assessor (KT) cross-checked the data extracted by the other assessors; where there was disagreement between assessors, each item was further discussed until consensus was reached.

Characteristics of the evaluation design

We categorised the primary evaluation design as either individual-level Randomised Controlled Trial (RCT), cluster RCT, case-control, interrupted time series, cohort study, before/after, cross-sectional, qualitative only, or multiple approaches (use of multiple evaluation designs to assess different project components). Studies were included in one of these categories based on standard definitions used for these type of study design,[8] and by using information provided in the methods section of the paper describing that evaluation. If inadequate information was provided to fit a design into a particular category, it was listed as 'other'. We categorised whether evaluations used a mixed-methods approach (analysing quantitative and qualitative data) versus analysing quantitative or qualitative data only.

We also assessed whether the primary outcome measured in the study was a health or a health risk factor outcome versus a health service delivery or process outcome (i.e. assessing utilisation, reach, satisfaction, quality, and implementation).[2] The type of organisation that led the evaluation, based on the first author's affiliation, was categorised as research institution/university, government, health service, NGO or commercial consultancy provider.

Analysis

Data were entered into a spreadsheet and imported to Stata version 14.1 for descriptive analysis.

Results

Our search identified 650 peer-reviewed documents; 145 full manuscripts were reviewed, of which 118 studies met the selection criteria, and described 109 unique interventions (Figure 1). There were eight included interventions on which multiple (two or three each) evaluations were published.

Evaluation design: 82.2% of studies included a quantitative component, with 49.2% utilising quantitative data only and 33.1% using both quantitative and qualitative data. The most common evaluation design was a before/after comparison (30.5%, n=36/118). Nine of 118 studies (7.6%) used an experimental design; this included 6 RCTs (5.1%) and three cluster-RCTs (2.5%). One of the nine studies (11.1%) using an experimental design also incorporated qualitative data; the remaining eight experimental studies (88.9%) analysed quantitative data only. Among the 109 studies that used a non-experimental design, 45.9% (n=50/109) analysed quantitative data only, 34.9% (n=38/109) analysed both quantitative and qualitative data, and 19.3% analysed qualitative data only (n=21/109).

Evaluation implementer: The majority of evaluations (72.0%, 85/118) were led by a university or research institution.

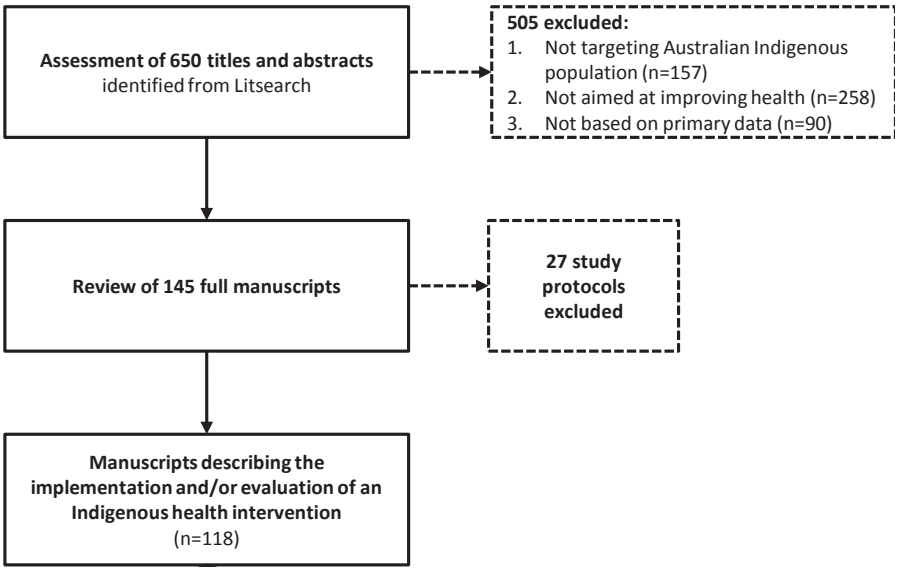
Type of outcome assessed: More than half (56.8%, n=67/118) of evaluations measured a health service delivery or process outcome; 33.1% (n=39/118) measured a health or health risk factor outcome, and 10.2% (n=12/118) measured both. All nine studies that employed an experimental design measured a health outcome or health risk factor, rather than a health service delivery or process outcome only. Of non-experimental evaluations, 61.5% (n=67/109) measured a health service delivery or process outcome, 28.4% (n=31/109) measured a health or health risk factor outcome, and 10.1% (n=11/109) measured both.

Discussion

Our search of the peer-reviewed literature identified relatively few program evaluations. We cannot make inferences on the proportion of all programs that are evaluated, but our results are consistent with a recent review identifying that only 10% of current Indigenous programs had been evaluated.[9] Our results also indicate that of the few published evaluations, the majority did not use optimal designs.

There have been increasing calls from policy makers for evidence from RCTs, which are considered the 'gold standard' for generating quantitative experimental evidence in health,[10,11] to inform policy and program decisions. Despite this we identified very few evaluations based on RCTs. One reason for this may have been that many interventions assessed were allocated by cluster (community or health service). In such cases individual RCTs are not possible, but experimental design such as cluster RCTs, stepped wedge or multiple baseline designs may be appropriate. We also identified very few of these designs. Experimental designs such as these are only possible when considered during program development and implementation, and if relevant to and supported by participating communities. Such evaluations may not be appropriate or required in every setting, particularly in the case of complex public health interventions or when the evaluation is conducted with a

Figure 1: PRISMA flow diagram.



small sample or in one setting, but as they provide a high level of evidence, they must be given due consideration during program design.[12] Calls for their use must also be based on an assessment of their utility, feasibility, and value to service agencies and communities during program planning. Inadequate resources and expertise within the organisations implementing and evaluating Indigenous health programs are likely to also pose a barrier to the use of experimental evaluation design.

Quantitative data provide a measure of the impact of health programs; qualitative data enable participative and collaborative evaluation and are valuable to contextualise and provide culturally relevant inferences about a program's utility and impact. Therefore, mixed methods approaches, utilising both quantitative and qualitative data, are likely to be the best approach to conducting complex program evaluations.[13,14] However, only a third of identified studies utilised a mixed qualitative/quantitative design. Stronger collaborations between qualitative and quantitative evaluators are needed, particularly by researchers using experimental evaluation designs. Only one of the nine studies using experimental design also included qualitative data.

Most of the published evaluations identified in our search were led by research institutions. This is consistent with the fact that researchers are incentivised to publish in peer-reviewed outlets. We identified only one evaluation led by a commercial contractor published in the peer-reviewed literature. External independent review of evaluations through processes such as peer-reviewed publication is likely to improve evaluation quality and facilitate uptake by researchers and policymakers. It would, however, also be likely to increase the costs and time required for evaluation and may require extra resourcing. The peer-review process should consider the quality of the evidence, as well as the appropriateness of the evaluation design. When no evidence is available for a program it may be more appropriate to first publish a pre/post evaluation in one setting to provide some evidence of effectiveness before conducting an expensive large-scale randomised trial.

Our findings underline the clear need for more high-quality evaluations of Indigenous health programs, including studies utilising experimental design alongside qualitative research that explores contextual barriers and enablers to successful implementation. However, the appropriateness of evaluation design must be considered in the context of the setting and the evidence base. Ensuring evaluations are a routine part of program planning can be achieved by working in partnership with non-research institutions, in particular Indigenous community organisations and local service providers, during program planning to increase their capacity to lead and implement such evaluations. This will require appropriate recognition and resourcing by those agencies commissioning and funding programs. By doing so, we will ensure such evaluations are of the highest possible quality and relevant to the needs of Indigenous communities.

Acknowledgements

This research received support from the Australian Primary Health Care Research Institute, Australian National University.

References

1. O'Donoghue L, Percival N, Laycock A, McCalman J, Tsey K, Armit C, et al. Evaluating Aboriginal and Torres Strait Islander health promotion activities using audit and feedback. *Aust J Prim Health*. 2014;20(4):339-44.
2. McCalman J, Tsey K, Clifford A, Earles W, Shakeshaft A, Bainbridge R. Applying what works: A systematic search of the transfer and implementation of promising Indigenous Australian health services and programs. *BMC Public Health*. 2012;12(1):600.
3. Sladek RM, Tieman JJ, Tyndall J, Phillips PA. Searching MEDLINE for Aboriginal and Torres Strait Islander health literature: Questionable sensitivity. *Health Info Libr J*. 2013;30(2):138-48.
4. Tieman JJ, Lawrence MA, Damarell RA, Sladek RM, Nikolof A. LIt. search: Fast tracking access to Aboriginal and Torres Strait Islander health literature. *Aust Health Rev*. 2014;38(5):541-5.
5. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337:a1655.
6. Brown CA, Lilford RJ. The stepped wedge trial design: A systematic review. *BMC Med Res Methodol*. 2006.6(1):54.
7. Hemming K, Girling A, Haines T, Lilford R. *Protocol: Consort Extension to stepped Wedge Cluster Randomised Controlled Trial*. Oxford (UK): University of Oxford The EQUATOR Network; 2014.
8. Effective Practice and Organisation of Care (EPOC). *What study designs should be included in an EPOC review?* EPOC Resources for review authors [cited 2017 Mar 31]. Available from: http://epoc.cochrane.org/epoc-specific-resources-review-authors Oslo: Norwegian Knowledge Centre for the Health Services; 2013.
9. Hudson S. *Mapping the Indigenous Program and Funding Maze*. Sydney (AUST): The Centre for Independent Studies; 2016.
10. Nelson JL. *Program Evalution: Are We Ready for RCTs*. London (UK): Active Learning Network for Accountability and Performance; 2008.
11. Goldacre B. *Building Evidence into Education*. London (UK): London School of Hygiene and Tropical Medicine; 2013.
12. Marley JV, Atkinson D, Kitaura T, Nelson C, Gray D, Metcalf S, et al. The Be Our Ally Beat Smoking (BOABS) study, a randomised controlled trial of an intensive smoking cessation intervention in a remote aboriginal Australian health care setting. *BMC Public Health*. 2014;14(1):1.
13. Howarth E, Devers K, Moore G, O'Cathain A, Dixon-Woods M. Contextual issues and qualitative research. In: Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, et al, editors. *Challenges, Solutions and Future Directions in the Evaluation of Service Innovations in Health Care and Public Health*. Southampton (UK): Health Services and Delivery Research; 2016. vol 4. p. 105-20.
14. Rossingh B, Yunupingu Y. Evaluating as an outsider or an insider: A two-way approach guided by the knowers of culture. *Eval J Australas*. 2016;6(3):5-14.